Obesity Levels of Healthy Cities

Kimberly Cable

Belleview University

DSC 530: Data Exploration and Analysis

Dr. Shankar Parajulee

June 4, 2022

Obesity Levels of Healthy Cities?

A study was done gathering data on all the cities in the world to see which were the top healthiest cities. But what makes these cities healthy? In my analysis, I wanted to know what were the determining factors that made them the top healthiest in the world. Specifically looking at obesity levels, the number of hours of sunshine a city gets per day, the cost of a bottle of water in the city’s country, the number of outdoor activities a city has, and the average number of hours a person worked.

I did some preliminary analysis to find out all the outliers in these variables and determine their descriptive statistics like mean, median, mode, and quartiles. Nothing looked out of the ordinary, so I looked at how the number of hours of sunshine affected obesity levels in healthy cities. I determined the most frequent number of hours of sunshine was five so I divided the data into two groups. First those cities with five hours of sunshine and the rest of the cities. With this, I looked at the obesity levels for each group. I found that with just five hours of sunshine the obesity levels were much lower than in the rest of the cities not having an average of five hours of sunshine per day.

Next, I looked at how each city’s cost of a bottle of water compared. Doing a cumulative distribution function, I found that at $1.56, 50% of the cities were below this cost. At $2.10, 75% of the cities were below this cost. At $0.75, only 25% of the cities were below this cost. It’s hard to determine if these costs were relevant as each city’s cost of living is different.

Comparing obesity levels with the number of hours of sunshine per day with a scatterplot, suggested that there is a positive correlation between the number of hours of sunshine and obesity levels. Pearson’s correlation coefficient of 0.31 with a p-value of 0.05 and an r-squared of 0.09 confirms there is no correlation between the number of hours of sunshine and obesity levels in a healthy city. Looking at these variables differently, I looked at the correlation between the cost of a bottle of water and the number of outdoor activities a healthy city has. The scatter plot indicates as the cost of a bottle of water goes up the number of outdoor activities also goes down. The Pearson’s coefficient is -0.29 with a p-value of 0.06 and an R-squared of 0.086. This also indicates no correlation between the cost of a bottle of water and the number of outdoor activities a city has.

Finally, I did some regression models on how the individual variable influenced obesity levels and then did a multiple regression with all of them to see if together they influenced obesity levels in a healthy city. I found that individually, they did not influence the obesity levels but together they did influence the obesity levels but a significant amount.

Looking at the analysis I performed, I feel I didn’t get very far in determining what obesity levels were in healthy cities. I think having data from non-healthy cities may have helped in the analysis. One assumption I made was that healthy cities had lower obesity levels which in my mind made them healthy which skewed my thinking and choosing the variables from the dataset that I did. With more research and better data, I think determining the obesity levels of healthy cities may come from much more variables and data.